Elect. Comm. in Probab. **13** (2008), 121–130

ELECTRONIC COMMUNICATIONS in PROBABILITY

ENTROPY ESTIMATE FOR K-MONOTONE FUNCTIONS VIA SMALL BALL PROBABILITY OF INTEGRATED BROWNIAN MOTION

FUCHANG GAO[1]
Department of Mathematics, P.O. Box 441103, University of Idaho, Moscow, ID 83844-1103
email: `fuchang@uidaho.edu`

Submitted August 1, 2007, *accepted in final form* February 13, 2008

AMS 2000 Subject classification: 46B50 (60G15, 62G07)
Keywords: metric entropy, k-monotone function, small ball probability, k-times integrated Brownian motion

Abstract
Metric entropy of the class of probability distribution functions on $[0, 1]$ with a k-monotone density is studied through its connection with the small ball probability of k-times integrated Brownian motions.

1 Introduction

In statistical consultation, one is often confronted with the problem that a client shows a graph of a certain observed frequency distribution and asks, "What theoretical probability distribution would fit this observed distribution?" This question becomes mathematically meaningful once one specifies the family of densities to consider and the distance to measure the deviation between the real density and the estimator of the density (Groeneboom (1985)). To answer such a question, non-parametric estimators, such as the Maximum Likelihood Estimator, are often used. It is well known that the rate of convergence of an estimator depends on the richness of the function class. In particular, it depends on the metric entropy of the function class. Thus, it is of interest to study the metric entropy of various shape-constrained function classes that have statistical significance.
For a bounded set T in a metric space equipped with distance d, the metric ε-entropy of T is defined as the logarithm of the minimum covering number, i.e, $\log N(\varepsilon, T, d)$, where

$$N(\varepsilon, T, d) := \min\left\{ m : \text{there exist } t_1, t_2, \dots, t_m \text{ such that } T \subset \bigcup_{k=1}^{m} \{t \in T : d(t, t_i) < \varepsilon\} \right\}.$$

Metric entropy was first introduced by A. N. Kolmogorov and has been extensively studied and applied in approximation theory, geometric functional analysis, probability theory, and

[1]RESEARCH SUPPORTED BY NSF GRANT DMS-0405855

complexity theory, etc.; e.g., see the books by Kolmogorov and Tihomirov (1961), Lorentz (1966), Carl and Stephani (1990), Edmunds and Triebel (1996). Among many beautiful results are the duality theorem (Tomczak-Jaegermann (1987), Artstein et.al (2004)), and the small ball probability connection (Kuelbs and Li (1993), Li and Linde (1999)), which will be used in this paper. Nevertheless, the estimate of metric entropy for specific function classes remains difficult, especially the lower bound estimate, which often requires a construction of a well-separated subset.

In this paper, we study the metric entropy estimate of a class of shape-constrained functions called k-monotone functions. k-monotone functions have been studied since at least the 1950s; for example, Williamson (1956) gave a characterization of k-monotone functions on $(0,\infty)$ in 1956. In recent years, there has been a lot of interest in statistics regarding this class of functions. We refer the recent paper by Balabdaoui and Wellner (2004) and the references therein for recent results and their statistical applications.

A function on a bounded interval, say $[0,1]$, is said to be m-monotone if $(-1)^k f^{(k)}(x)$ is non-negative, non-increasing, and convex for $0 \le k \le m-2$ if $m \ge 2$, and $f(x)$ is non-negative, non-increasing if $m=1$. Let us note that in dealing with the metric entropy of this function class under L^p norms, $1 \le p < \infty$, we can always assume that the functions are differentiable infinitely many times by using the the following basic lemma.

Lemma 1.1. *m-monotone C^∞ functions are dense in m-monotone functions under L^p norm, $1 \le p < \infty$.*

Proof. The idea of the proof is simple. Basically, we can approximate a continuous function f by $f * K$ for some C^∞ kernel K without changing the m-monotonicity. However, this requires an extension of f to a larger interval containing $[0,1]$ while maintaining the m-monotonicity, which is not immediately clear for $m \ge 2$. Thus, we give a detailed proof for the case $m \ge 2$.

If f is m-monotone on $[0,1]$ for $m \ge 2$, then, by definition, $(-1)^{m-2} f^{(m-2)}$ is non-negative increasing and convex. For any $\varepsilon > 0$, we can find a piecewise linear non-negative increasing convex function g_{m-2}, such that $\|(-1)^{m-2} f^{(m-2)} - g_{m-2}\|_p < \varepsilon$. Extend g_{m-2} to $\mathbb{R}$, so that g_{m-2} is supported on $[-1,1]$, and once restricted on $[-1,1]$, g_{m-2} is a continuous non-negative increasing convex function. Let K_ε be a $C^\infty(-\infty,\infty)$ kernel supported on $[0,1]$ such that

$$\|g_{m-2} - h_{m-2}\|_{L^p[0,1]} \le \varepsilon,$$

where $h_{m-2} = g_{m-2} * K_\varepsilon$. Clearly $h_{m-2} \in C^\infty$. Because for each fixed $x \in \operatorname{supp}(K_\varepsilon) \subset [0,1]$, $g_{m-2}(t-x)$ is a non-negative, increasing and convex function of t, and

$$h_{m-2}(t) = \int_{-\infty}^{\infty} g_{m-2}(t-x) K_\varepsilon(x)\,dx = \int_0^1 g_{m-2}(t-x) K_\varepsilon(x)\,dx,$$

we see that h_{m-2} is also a non-negative, increasing and convex function on $[0,1]$.

Now, define

$$h_{m-3}(t) := (-1)^{m-3} f^{(m-3)}(1) + \int_t^1 h_{m-2}(x)\,dx.$$

Because

$$(-1)^{m-3} f^{(m-3)}(t) - h_{m-3}(t) = \int_t^1 \left[(-1)^{m-2} f^{(m-2)}(x) - h_{m-2}(x)\right] dx,$$

we have

$$\left\|(-1)^{m-3} f^{(m-3)} - h_{m-3}\right\|_p \le \left\|(-1)^{m-2} f^{(m-2)} - h_{m-2}\right\|_p \le 2\varepsilon.$$

Repeating this process, we can obtain an m-monotone C^∞ function h_0 such that $\|f - h_0\|_p \leq 2\varepsilon$. $\square$

In view of Lemma 1.1, we will simply say for convenience that a function is m-monotone if $(-1)^k f^{(k)}(x) \geq 0$ for all $0 \leq k \leq m$.
Our main result is the following

Theorem 1.2. Let $\mathcal{M}_m$ be the class of probability distribution functions on $[0,1]$ with an m-monotone density function. Then

$$\log N(\varepsilon, \mathcal{M}_m, \|\cdot\|_2) \asymp \varepsilon^{-1/(m+1)}.$$

where $a \asymp b$ means $a = O(b)$ and $b = O(a)$ as $\varepsilon \to 0^+$.

This is a generalization of a result due to Van de Geer (1991) based on the earlier work of Birman and Solomjak (1967) (see also Van der Vaart and Wellner (1996); Theorem 2.7.5) which dealt with the case $m = 0$. The elegance of this paper is the method, which reveals a precise connection with the small deviation probability of m-times integrated Brownian motions. Blei at. al (2007) contains another application of this method.)

2 A Characterization

First, we need a characterization of the function class $\mathcal{M}_m$. Recall that Williamson (1956) proved that a function g is k-monotone on $(0,\infty)$ if and only if there exists a non-decreasing function γ bounded at 0 such that

$$g(x) = \int_0^\infty (1-tx)_+^{k-1} d\gamma(t), \quad x > 0$$

where $y_+ = y1_{(0,\infty)}(y)$. The following theorem gives a similar characterization for the function class $\mathcal{M}_m$.

Theorem 2.1. A function $F(x)$ is a probability distribution function on $[0,L]$ with a m-monotone density if and only if it is of the form

$$F(x) = 1 - \left[a_1(L-x) + a_2(L-x)^2 + \cdots + a_m(L-x)^m + \int_x^L \left(1 - \frac{x}{t}\right)^m d\mu(t)\right], \qquad (1)$$

where $a_1, a_2, ..., a_m \geq 0$, μ is a non-negative measure on $[0,L]$, and

$$a_1 L + a_2 L^2 + \cdots + a_m L^m + \|\mu\| = 1.$$

Proof. Suppose F is a probability distribution function on $[0,L]$ with an m-monotone density. Then

$$(-1)^m F^{(m)}(t) = (-1)^m f^{(m-1)}(t)$$

is non-decreasing. Thus, $d\mu(t) := \frac{(-1)^m t^m}{m!} dF^{(m)}(t)$ defines a non-negative measure on $[0, L]$. Repeatedly using integration by parts gives

$$\begin{aligned}\int_x^L \left(1 - \frac{x}{t}\right)^m d\mu(t) &= \int_x^L \frac{(x-t)^m}{m!} dF^{(m)}(t) \\ &= \sum_{k=1}^m \frac{F^{(k)}(L)}{k!}(x-L)^k + F(L) - F(x) \\ &= \sum_{k=1}^m \frac{(-1)^k F^{(k)}(L)}{k!}(L-x)^k + 1 - F(x).\end{aligned}$$

Let $a_k = \frac{(-1)^{k-1} F^{(k)}(L)}{k!}$. Then $a_k \geq 0$, and we have

$$F(x) = 1 - \left[a_1(L-x) + a_2(L-x)^2 + \cdots + a_m(L-x)^m + \int_x^L \left(1 - \frac{x}{t}\right)^m d\mu(t)\right].$$

It remains to prove that $a_1 L + a_2 L^2 + \cdots + a_m L^m + \|\mu\| = 1$. Note that by repeatedly using integration by parts, we also have

$$\begin{aligned}\|\mu\| &= \int_0^L \frac{(-1)^m t^m}{m!} dF^{(m)}(t) \\ &= F(L) - F(0) - \sum_{k=1}^m (-1)^{k-1} \frac{F^{(k)}(L)}{k!} L^k \\ &= 1 - (a_1 L + a_2 L^2 + \cdots + a_m L^m).\end{aligned}$$

This proves that F can be expressed as (1). The other direction is trivial. □

The proof of Theorem 2.1 also gives the following

Corollary 2.2. A function f is an integrable m-monotone function on $[0, L]$ if and only if it can be expressed as

$$f(x) = a_1 + a_2(L-x) + \cdots + a_m(L-x)^{m-1} + \int_x^L \frac{m}{t}\left(1 - \frac{x}{t}\right)^{m-1} d\mu(t),$$

where $a_1, a_2, ..., a_m \geq 0$, and μ is a non-negative measure. Furthermore,

$$a_1 L + \frac{a_2}{2} L^2 + \frac{a_3}{3} L^3 + \cdots + \frac{a_m}{m} L^m + \|\mu\| = \int_0^L f(x) dx.$$

Remark 2.3. Corollary 2.2 is an extension of the following result of Williamson (See Balabdaoui and Wellner (2004); Lemma 2.1): a function g is an integrable m-monotone function on $(0, \infty)$ if and only if it is of the form

$$g(x) = \int_x^\infty \frac{m(t-x)^{m-1}}{t^m} d\mu(t),$$

where μ is a finite measure on $(0, \infty)$.

3 Proof of the Main Result

We denote by $\mathcal{Q}_m$ the class of functions on $[0,1]$ of the form

$$1-\int_x^1\left(1-\frac{x}{t}\right)^m d\mu(t),$$

where $\mu(t)$ is a non-negative measure with total variation bounded. We also denote by $\mathcal{P}_m$ the class of polynomials of the form $a_1(1-x)+\cdots+a_m(1-x)^m$, with $a_1,a_2,...,a_m\ge 0$ and $a_1+\cdots+a_m\le 1$. Then Theorem 2.1 implies that $\mathcal{Q}_m\subset\mathcal{M}_m\subset\mathcal{Q}_m+\mathcal{P}_m$. Thus,

$$N(\varepsilon,\mathcal{Q}_m,\|\cdot\|_2)\le N(\varepsilon,\mathcal{M}_m,\|\cdot\|_2)\le N(\varepsilon/2,\mathcal{Q}_m,\|\cdot\|_2)N(\varepsilon/2,\mathcal{P}_m,\|\cdot\|_2). \tag{2}$$

On the other hand, it is easy to see that

$$N(\varepsilon,\mathcal{P}_m,\|\cdot\|_2)\le\left(1+\frac{m}{\varepsilon}\right)^m. \tag{3}$$

Indeed, the set

$$\{a_1(1-x)+\cdots+a_m(1-x)^m : a_i\in\{1/N,2/N,...,N/N\},1\le i\le m\}$$

forms an m/N-net of $\mathcal{P}_m$, and there are only N^m elements in this set. By choosing $N=\lceil m/\varepsilon\rceil$, inequality (3) follows. Substituting (3) into (2), we obtain

$$\log N(\varepsilon,\mathcal{M}_m,\|\cdot\|_2)\asymp\log N(\varepsilon,\mathcal{Q}_m,\|\cdot\|_2),$$

provided that we show $\log N(\varepsilon,\mathcal{Q}_m,\|\cdot\|_2)\asymp\varepsilon^{-\alpha}$ for some $\alpha>0$.

To estimate the covering number $N(\varepsilon,\mathcal{Q}_m,\|\cdot\|_2)$, we introduce an auxiliary function class $\widetilde{\mathcal{Q}}_m$ that consists of all the functions on $[0,1]$ that can be expressed as $1-\int_x^1(1-x/t)^m d\nu(t)$, where ν is a signed measure on $[0,1]$ with total variation bounded by 1. The benefit of using this auxiliary function class is that $\widetilde{\mathcal{Q}}_m$ has a certain useful symmetry, which will become clear later in the proof.

It is clear that $\mathcal{Q}_m\subset\widetilde{\mathcal{Q}}_m$. So, $N(\varepsilon,\mathcal{Q}_m,\|\cdot\|_2)\le N(\varepsilon,\widetilde{\mathcal{Q}}_m,\|\cdot\|_2)$. On the other hand, if $F\in\widetilde{\mathcal{Q}}_m$, then there exists a signed measure ν with total variation bounded by 1, such that

$$F(x)=1-\int_x^1\left(1-\frac{x}{t}\right)^m d\nu(t).$$

Let $\mu_1:=\nu_+$ and $\mu_2:=\nu_-$. We have

$$\begin{aligned}F(x) &= 1-\int_x^1\left(1-\frac{x}{t}\right)^m d\nu(t)\\ &= \left[1-\int_x^1\left(1-\frac{x}{t}\right)^m d\mu_1(t)\right]-\left[1-\int_x^1\left(1-\frac{x}{t}\right)^m d\mu_2(t)\right]+1.\end{aligned}$$

This means that for any $F\in\widetilde{\mathcal{Q}}_m$, there exist $F_1,F_2\in\mathcal{Q}_m$ such that $F(x)=F_1(x)-F_2(x)+1$ for all $x\in[0,1]$, or $\widetilde{\mathcal{Q}}_m\subset\mathcal{Q}_m-\mathcal{Q}_m+1$. This immediately implies that

$$N(\varepsilon,\widetilde{\mathcal{Q}}_m,\|\cdot\|_2)\le N(\varepsilon/2,\mathcal{Q}_m,\|\cdot\|_2)^2.$$

Hence,

$$\log N(\varepsilon, \mathcal{Q}_m, \|\cdot\|_2) \asymp \log N(\varepsilon, \widetilde{\mathcal{Q}}_m, \|\cdot\|_2),$$

provided that $\log N(\varepsilon, \widetilde{\mathcal{Q}}_m, \|\cdot\|_2)$ is of the order $\varepsilon^{-\alpha}$ for some $\alpha > 0$, which will be proved later.
To estimate $\log N(\varepsilon, \widetilde{\mathcal{Q}}_m, \|\cdot\|_2)$, we notice that for any two functions $F_1, F_2 \in L^2[0,1]$, by Parseval's identity,

$$\|F_1 - F_2\|_2^2 = \sum_{n=1}^{\infty} \langle F_1 - F_2, \phi_n \rangle^2 ,$$

where $\{\phi_n\}_{n=1}^{\infty}$ is an orthonormal basis of $L^2[0,1]$. Thus, the covering number $N(\varepsilon, \widetilde{\mathcal{Q}}_m, \|\cdot\|_2)$ is the same as the covering number $N(\varepsilon, S, \|\cdot\|_{l_2})$, where

$$S = \{(a_1, a_2, ...) : a_n = \langle F, \phi_n \rangle , n \in \mathbb{N}, F \in \widetilde{\mathcal{Q}}_m\}.$$

Of course, $N(\varepsilon, S, \|\cdot\|_{l_2})$ is the same as $N(\varepsilon, T, \|\cdot\|_{l_2})$, where

$$T = \left\{ (a_1, a_2, ...) : a_n = \langle F, \phi_n \rangle , n \in \mathbb{N}, F(x) = \int_x^1 \left(1 - \frac{x}{t}\right)^m d\mu(t), \|\mu\| \leq 1 \right\}.$$

Note that T is a symmetric convex subset of l^2. (The purpose of introducing the auxiliary function class $\widetilde{\mathcal{Q}}_m$ is to create this symmetry.)
By the duality theorem of metric entropy (Tomczak-Jaegermann (1987), Artstein et.al (2004)), provided that either side of the relation below is of the order $\varepsilon^{-\alpha}$ for some $\alpha > 0$,

$$\log N(\varepsilon, T, \|\cdot\|_{l_2}) \asymp \log N(\varepsilon, D_2, \|\cdot\|_{T^\circ}), \tag{4}$$

where D_2 is the unit ball of l^2 and $\|\cdot\|_{T^\circ}$ is a norm induced by the set

$$T^\circ = \left\{ (x_1, x_2, ...) \in \mathbb{R}^\infty : \sup_{(t_1, t_2, ...) \in T} \left| \sum_{i=1}^{\infty} x_i t_i \right| \leq 1 \right\}.$$

Now, let us take a closer look at the set T. Note that by changing the order of integration, we can write

$$\begin{aligned} \langle F, \phi_n \rangle &= \int_0^1 \int_x^1 \left(1 - \frac{x}{t}\right)^m d\mu(t) \ \phi_n(x) dx \\ &= \int_0^1 \int_0^t \left(1 - \frac{x}{t}\right)^m \phi_n(x) dx \ d\mu(t). \end{aligned}$$

Thus, T is the absolute convex hull of the set

$$\left\{ (a_1(t), a_2(t), ...) : a_n(t) = \int_0^t \left(1 - \frac{x}{t}\right)^m \phi_n(x) dx, n = 1, 2, ...; t \in [0,1] \right\}.$$

Next, we relate T to an m-times integrated Brownian motion. Let $W(t)$, $t \in [0,1]$ be the Brownian motion on $[0,1]$. Writing $W(t)$ in a canonical expansion, we have

$$W(t) \stackrel{\mathrm{d}}{=} \sum_{n=1}^{\infty} \int_0^t \phi_n(x) dx \cdot \xi_n, \tag{5}$$

where ξ_n are independent $N(0,1)$ random variables. Let B_m be an m-times integrated Brownian motion, i.e.,

$$B_m(t) = \int_0^t \int_0^{x_1} \int_0^{x_2} \cdots \int_0^{x_{n-1}} W(x_n) dx_n \cdots dx_3 dx_2 dx_1.$$

By using the canonical expansion (5) and changing the order of integration, we have

$$\begin{aligned} B_m(t) &= \int_0^t (t-x)^m dW(x) \\ &= \sum_{n=1}^{\infty} \int_0^t (t-x)^m \phi_n(x) dx \cdot \xi_n. \end{aligned}$$

Thus,

$$\frac{B_m(t)}{t^m} = \sum_{n=1}^{\infty} \int_0^t \left(1 - \frac{x}{t}\right)^m \phi_n(x) dx \cdot \xi_n.$$

Note that the right hand side is exactly the inner product of the vector $(\xi_1, \xi_2, ...)$ and a vector in the set T. Thus,

$$T^\circ = \left\{ (\xi_1, \xi_2, ...) \in \mathbb{R}^\infty : \sup_{t \in [0,1]} \left| \frac{B_m(t)}{t^m} \right| \le 1 \right\}.$$

We will use the connection between metric entropy and small ball probability to estimate $\log N(\varepsilon, D_2, \|\cdot\|_{T^\circ})$. By a general connection between small ball probability and metric entropy discovered by Kuelbs and Li (1993) and completed in Li and Linde (1999), the covering number $N(\varepsilon, D_2, \|\cdot\|_{T^\circ})$ is connected with the Gaussian measure of $\{x \in \mathbb{R}^\infty : \|x\|_{T^\circ} \le \varepsilon\}$, that is, the small ball probability $\mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon\}$. The precise connection is as follows:

$$\log N(\varepsilon, D_2, \|\cdot\|_{T^\circ}) \asymp \varepsilon^{-\alpha} \quad \text{if and only if} \quad \log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon) \asymp -\varepsilon^{-\frac{2\alpha}{2-\alpha}}. \tag{6}$$

Therefore, it remains to estimate $\log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon)$.
It is clear that

$$\log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon) \le \log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)| \le \varepsilon). \tag{7}$$

On the other hand, by the Weak Gaussian Correlation Inequality (Li (1999), Schechtman et.al (1998)) and the scaling property of m-times integrated Brownian motion that $B_m(ct) \overset{\mathrm{d}}{=} c^{m+1/2} B_m(t)$, we have for any $0 < \lambda < 1$

$$\begin{aligned} \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon) &\ge \mathbb{P}(\sup_{t\in[0,\delta]} |B_m(t)/t^m| \le \lambda\varepsilon) \cdot \mathbb{P}(\sup_{t\in[\delta,1]} |B_m(t)/t^m| \le \sqrt{1-\lambda^2}\varepsilon) \\ &= \mathbb{P}(\sup_{s\in[0,1]} |B_m(s)/s^m| \le \delta^{-1/2}\lambda\varepsilon) \cdot \mathbb{P}(\sup_{t\in[\delta,1]} |B_m(t)/t^m| \le \sqrt{1-\lambda^2}\varepsilon) \\ &\ge \mathbb{P}(\sup_{s\in[0,1]} |B_m(s)/s^m| \le \delta^{-1/2}\lambda\varepsilon) \cdot \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)| \le \delta^m \sqrt{1-\lambda^2}\varepsilon). \end{aligned}$$

By choosing $\delta = \frac{m}{4m+2}$ and $\lambda = 2\sqrt{\delta}$, we have

$$\mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon) \ge \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le 2\varepsilon) \cdot \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)| \le \frac{m^m}{2^m(2m+1)^{m+1/2}}\varepsilon).$$

By iteration, we have

$$\log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon) \ge C \log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)| \le c\varepsilon)$$

for some constants $C > 0$ and $0 < c < 1$, which, together with (7), implies

$$\log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)| \le \varepsilon) \asymp \log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)/t^m| \le \varepsilon), \tag{8}$$

provided that the right-hand-side is of the order $-\varepsilon^{-\beta}$ for some $\beta > 0$.
However, it was proved in Chen and Li (1999) that

$$-\log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)| \le \varepsilon) \asymp \varepsilon^{-\frac{2}{2m+1}}. \tag{9}$$

Putting (4), (6), (8) and (9) together, we conclude that

$$\log N(\varepsilon, \mathcal{M}_m, \|\cdot\|_2) \asymp \varepsilon^{-\frac{1}{m+1}}.$$

4 Some Remarks

In statistical applications, one may also want to consider the metric entropy of m-monotone densities. That is, m-monotone functions on $[0,1]$ satisfying $\|g\|_1 = 1$ with $m \ge 1$. If we denote this class of functions by $\mathcal{D}_m$, then a similar argument gives

$$\log N(\varepsilon, \mathcal{D}_m, \|\cdot\|_2) \asymp \varepsilon^{-\frac{1}{m}}.$$

Also note that for the class of m-monotone functions on $[0,1]$, even if we only consider the functions with continuous $f^{(m)}$, we generally cannot assume $f^{(m)}$ to be bounded. One might think that by restricting $f^{(m)}$ to be bounded by a certain number, one would obtain a smaller metric entropy. However, through a similar argument one can show that the metric entropy of the subclass of m-monotone function on $[0,1]$ with $|f^{(k)}| \le 1$ for all $k \le m$ has order $\varepsilon^{-1/m}$ as well.
Let us also remark that instead of requiring $(-1)^k f^{(k)} \ge 0$ for all $1 \le k \le m$, one can require $(-1)^{\varepsilon_k} f^{(k)} \ge 0$ for all $1 \le k \le m$, where $\varepsilon_k \in \{0,1\}$. We call such a class of functions as a general m-monotone class. We note that not only the same result as Theorem 1.2 holds for that class, but also that the same argument works. Indeed, if in the definition of m-times integrated Brownian motion

$$B_m(t) = \int_0^t \int_0^{x_1} \int_0^{x_2} \cdots \int_0^{x_{n-1}} W(x_n) dx_n \cdots dx_3 dx_2 dx_1,$$

we replace some of the integral limits "from 0 to x_i" by "from x_i to 1", we obtain a general m-times integrated Brownian motion $\widetilde{B}_m$, which was introduced in Gao et.al (2003). By interchanging the order of integration, a general m-times integrated Brownian motion can then be expressed as $\int_0^1 K(t,s)dW(s)$ for some kernel $K(t,s)$. By properly choosing the integral limits (either from 0 to x_i, or from x_i to 1) in the definition of $\widetilde{B}_m$, we can make

$$(-1)^{\varepsilon_k}\frac{\partial^{(k)}}{\partial t^k}K(t,s) \le 0.$$

Denoting $Q(t) = \int_0^s K(t,s)ds$, one can characterize the class of probability distribution functions on $[0,1]$ with a general m-monotone density as in Theorem 2.1, and argue that the problem of estimating the metric entropy of the function class under the L^2 norm becomes the problem of estimating the small ball probability of $\widetilde{B}_m(t)/Q(t)$ under the supremum norm, which eventually leads to the problem of estimating the small ball probability of general m-times integrated Brownian motion. However, it was recently proved by Gao and Li (2006) that

$$\log \mathbb{P}(\sup_{t\in[0,1]} |\widetilde{B}_m(t)| \le \varepsilon) \asymp \log \mathbb{P}(\sup_{t\in[0,1]} |B_m(t)| \le \varepsilon),$$

for all general m-times integrated Brownian motions. Thus, we conclude that Theorem 1.1 continues to hold if $\mathcal{M}_m$ is replaced by the class of probability distribution functions with a general m-monotone density on $[0,1]$.

Acknowledgement: The author thanks Professor Jon Wellner for bringing the author's attention to the k-monotone class of functions and explaining its statistical significance. Thanks also go to the referee for many valuable suggestions.

References

[1] Artstein, S.; Milman, V.; Szarek, S. (2004). Duality of metric entropy. *Ann. of Math.* **159**, 1313–1328. MR2113023

[2] Balabdaoui, F.; Wellner, J. A. (2004). Estimation of a k.monotone density, part 1: characterizations, consistency, and minimax lower bounds. Technical Report 459, University of Washington Department of Statistics.

[3] Birman, M. Š.; Solomjak, M. Z. (1967) Piecewise polynomial approximations of functions of classes ${W_p}^{\alpha}$. (Russian) Mat. Sb. (N.S.) **73** no. 115, 331–355. MR0217487

[4] Blei, R.; Gao, F.; Li, W. V. Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc.* 135 (2007), no. 12, 4009–4018 MR2341952

[5] Chen, X.; Li, W. V. (2003). Quadratic functionals and small ball probabilities for the m-fold integrated Brownian motion. *Ann. Probab.* **31**, no. 2, 1052–1077. MR1964958

[6] Carl, B.; Stephani, I. (1990). *Entropy, Compactness and Approximation of Operators*, Cambridge University Press, Cambridge. MR1098497

[7] Edmunds, D. E.; Triebel, H. (1996). *Function Spaces, Entropy Numbers and Differential Operators*, Cambridge Univ. Press. Cambridge. MR1410258

[8] Gao, F.; Hannig, J.; Torcaso, F. (2003). Integrated Brownian motions and exact L_2-small balls. *Ann. Probab.* **31**, no. 3, 1320–1337. MR1989435

[9] Gao, F.; Li, W. V. (2006). Logarithmic level comparison for small deviation probabilities. *J. Theoret. Probab.* **19** no. 3, 535–556. MR2280509

[10] Groeneboom, P. (1985). Estimating a monotone density. Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II. Lucien M. LeCam and Richard A. Olshen eds. Wadsworth, New York. 529 - 555. MR0822052

[11] Kolmogorov, A.N.; Tihomirov, V.M. (1961). ε-entropy and ε-capacity of sets in function spaces, *Uspehi Mat. Nauk.* **14** (1959), 3-86. English transl. in *Amer. Math. Soc. Transl.* **17**, 277-364. MR0112032

[12] Kuelbs, J.; Li, W.V. (1993). Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.* **116**, 133-157. MR1237989

[13] Li, W. V. (1999) A Gaussian correlation inequality and its applications to small ball probabilities. *Electron. Comm.* Probab. **4**, 111–118 MR1741737

[14] Li, W.V.; Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27**, 1556-1578. MR1733160

[15] Lorentz, G. (1966). Metric entropy and approximation. *Bull. Amer. Math. Soc.* **72**, 903–937. MR0203320

[16] Schechtman, G.; Schlumprecht, Th.; Zinn, J. (1998). On the Gaussian measure of the intersection. *Ann. Probab.***26**, no. 1, 346–357. MR1617052

[17] Tomczak-Jaegermann, N. (1987). Dualité des nombres d'entropie pour des opérateurs à valeurs dans un espace de Hilbert. (French) *C. R. Acad. Sci. Paris Sér. I Math.* **305** , no. 7, 299–301.

[18] Van de Geer, S. (1991). The entropy bound for monotone functions. Report 91-10. University of Leiden.

[19] Van der Vaart, A.; Wellner, J. (1996). *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York. MR1385671

[20] Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.* **23**, 189–207. MR0077581